McKinsey & Company

Technology, Media & Telecommunications Practice

AI infrastructure: A new growth avenue for telco operators

Operators could provide the backbone for today's AI economy to reignite growth. But success will hinge on effectively navigating complex market dynamics, uncertain demand, and rising competition.

This article is a collaborative effort by Abhyuadaya Shrivastava, Brendan Gaffey, Gustav Grundin, Sebastian Cubela, and Tomás Lajous, with Akshat Agarwal, Dapo Orimoloye, Lorraine Salazar, Miguel Frade, and Nicholas Shaw, representing views from McKinsey's Technology, Media & Telecommunications Practice and QuantumBlack, AI by McKinsey.

© Getty Images

February 2025

Telecom operators have long provided the infrastructure to power communication and connect people. Now they are poised to take on a new role: building the AI infrastructure that enables enterprises, governments, and consumers to unlock AI's full potential.

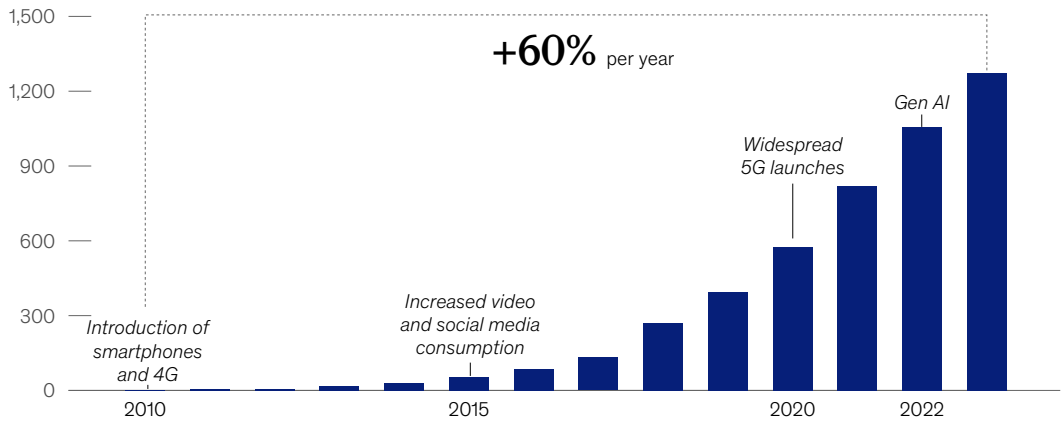
But finding the sweet spot to capture meaningful revenue and renewed success[1] will require speed and precision amid complex market dynamics, uncertain demand, and significant competitive headwinds. Delays or missteps may heighten operators' risk of falling further behind hyperscalers and other new market entrants, which have been the primary beneficiaries of growing data consumption over the last decade while telco revenues remained largely stagnant (Exhibit 1).

[1] Andrea Travasoni, Luca Fiandro, Venkat Atluri, and Marco Labianca, "The future of telcos: Mapping the routes to renewed success," McKinsey, December 18, 2024.

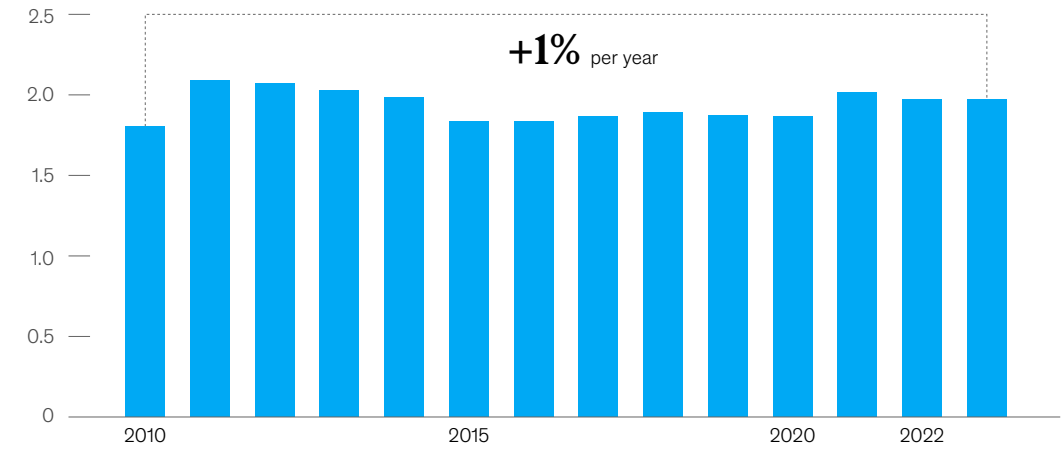
Exhibit 1

Telcos did not capture a fair share of growth from tech disruptions in the last decade.

Global mobile data traffic, exabytes



Telecom global revenue,[1] $ trillion



[1]Total fixed and mobile service revenues converted at the average 2023 $ exchange rates.
Source: Ericsson Mobility Visualizer; Omdia Communications Provider Revenue and Capex Forecast, Q2 2022, Q1 2024, Q2 2024

McKinsey & Company

Telcos that can move quickly may have an edge in the marketplace. Their vast footprint, geographical reach, and ability to manage large-scale networks and variable demand position them to meet the rising need for high-performance compute and connectivity driven by gen AI and agentic AI applications. By 2030, data center demand could more than triple[2] (Exhibit 2), with capacity expanding into new markets.[3]

Building and operating these data centers will require fiber to connect them (either to each other or to end users), space and power to host them, and GPUs (graphic processing units) for AI model training and inferencing. Enterprises will also need configurable network solutions—basically the next generation of software-defined networks (SDNs)—to efficiently manage requirements for different AI workloads running over the cloud. These can present opportunities for telcos, each varying in investment size, risk, and revenue potential. The viability of each opportunity for a given operator will vary based on regional demand, market structure, and the organization's asset base, appetite for risk, and financial position.

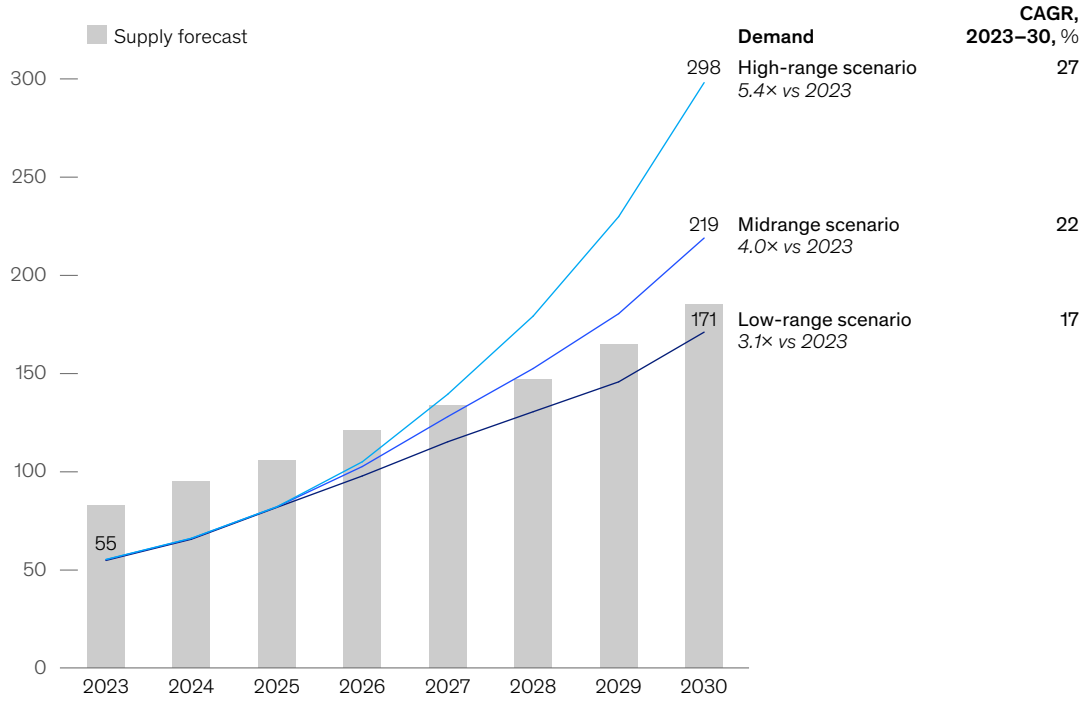
Unsurprisingly, competition in this space is intensifying from many quarters—cloud service providers, especially hyperscalers, cloud exchange providers, and other new market entrants. Adding to this, the emergence of more cost-efficient AI models may accelerate demand

[2] "AI power: Expanding data center capacity to meet growing demand," McKinsey, October 29, 2024.
[3] Example locations: Leesburg, Virginia, and York, South Carolina (both in the United States); Huizhou, province of Guangdong, China; and Johor, Malaysia.

Exhibit 2

Based on AI adoption scenarios, global data center demand will more than triple to at least 170 gigawatts by 2030.

Estimated global data center supply and demand, gigawatts



Source: McKinsey proprietary data center demand model

McKinsey & Company

for AI inferencing and distributed AI compute,[4] so it is worthwhile for operators to evaluate their options now.

In this article, we share four main paths telcos could pursue—individually or in combination—and potential considerations that leaders can weigh to balance ambition with a clear-eyed understanding of the associated risk. We also lay out insights on the operating model shifts necessary for these investments. (Telco leaders may also wish to consider opportunities to monetize gen AI software solutions, models, and tooling, which we explore, along with additional insights, in "Scaling the AI-native telco".[5])

Pathways to growth: Exploring four strategic options

Achieving meaningful growth starts with understanding the trade-offs in each strategic option. Each path offers distinct advantages and challenges as leaders aim to reignite growth and drive differentiation.

Path 1: Connecting new data centers with fiber

Global colocation companies and hyperscalers have announced they will break ground on more than 2,600 new data centers. About a quarter of these new data centers will be in cities with no operational data centers today (Exhibit 3). By the early 2030s, colocation providers and hyperscalers are expected to operate nearly 11,000 data centers worldwide (Exhibit 4).

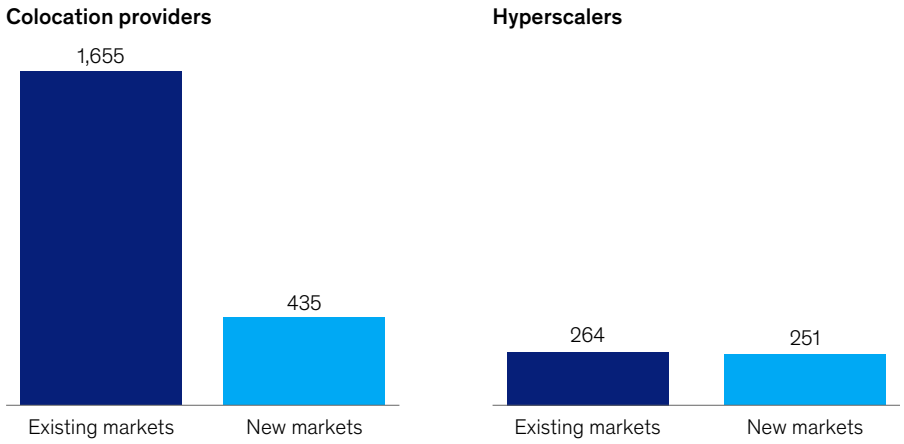
However, building data centers is just the start; colocation providers and hyperscalers must also connect the data centers via fiber to scale AI workloads for customers. In some cases, making these connections will be possible using existing fiber routes that meet customer specifications and

[4] Distributed AI compute enables processing of AI workloads at edge locations instead of centralized data centers.
[5] "Scaling the AI-native telco," McKinsey, February 27, 2025.

Exhibit 3

One-quarter of the new data centers announced by hyperscalers and colocation providers are in new markets.

New data centers announced,[1] global, number of facilities



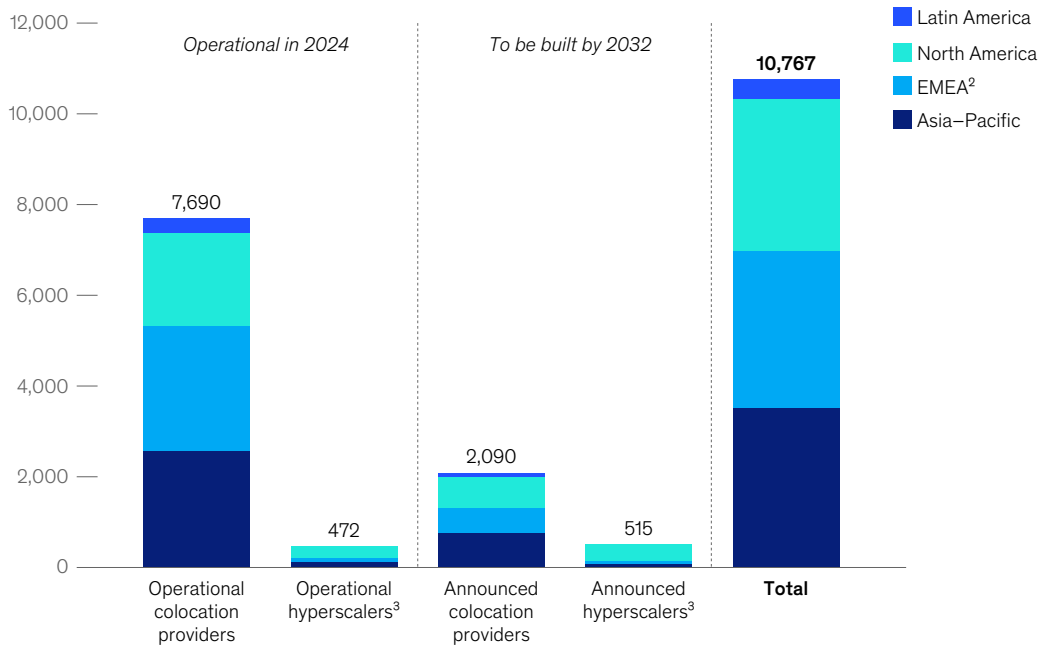
[1]84 announced facilities excluded as data on city location unavailable.
Source: 451 Research, part of S&P Global Market Intelligence (data as of Dec 2024)

McKinsey & Company

Exhibit 4

The number of data centers globally across colocation providers and hyperscalers is expected to reach approximately 11,000 by the early 2030s.

Data centers across colocation providers and hyperscalers,[1] number of facilities



[1]84 announced facilities excluded as data on city location unavailable.
[2]Europe, Middle East, and Africa.
[3]Latin America has 2 operational hyperscaler facilities in 2024 and 15 announced to be built by 2032.
Source: 451 Research, part of S&P Global Market Intelligence (data as of Dec 2024)

McKinsey & Company

project requirements. In others, a partial buildout may be required to connect new data centers.

Our research finds that this demand will create at least a $30 billion to $50 billion global market opportunity. While operators may have an initial edge, opportunities will vary by market. In many countries in Europe and Asia–Pacific, hyperscalers need a telco operator's license to lay fiber, giving existing operators priority in choosing which national and metro markets they will invest in and which they will lease to hyperscalers. In the United States, telcos may find that their existing footprint enables them to minimize permitting and construction work, especially when they can showcase unique routes.

At the same time, hyperscalers—telco's biggest competition in this space—may find it more practical to lease or buy existing fiber infrastructure from diverse telcos and other providers. Doing so enhances redundancy, helping protect against outages while also enabling them to scale network capacity faster. Prime examples include Verizon's agreements with Google Cloud and Meta to provide network infrastructure (and more) for their AI workloads.[6]

Because hyperscalers have vast data center networks, catering to their needs will be critical. This means telcos can expect a preference for dark fiber,[7] which allows customers to install and manage their networking equipment and optimize

[6] "Verizon unveils AI strategy to power next-gen AI demands," Verizon press release, January 24, 2025.
[7] Unused or "unlighted" fiber optic cable that is already deployed.

and scale bandwidth. Lumen, for instance, inked a fiber deal with Microsoft[8] that includes providing access to dark fiber. Wavelength services on lit fiber[9] may be desirable for temporary and backup lines and tertiary links in markets with a limited supply of dark fiber.

Path 2: Enabling high-performance cloud access with intelligent network services

As enterprises run AI workloads on the cloud and their business requirements become increasingly complex, they will need intelligent network services that give them more flexibility and control in managing the network. These services can help enterprises rein in cloud data transfer fees (commonly known as egress costs), which are estimated to exceed $70 billion to $80 billion annually.[10] An organization might use intelligent network services to dynamically route certain gen AI workloads within specific countries to meet regulatory or national security requirements or monitor workloads to more accurately predict and reduce egress costs. Such capabilities can also allow gen AI apps to automatically request access to additional network capability for latency-sensitive workloads, such as AI inferencing, which McKinsey analysis projects to account for a majority of AI workloads by 2030 (up from just 15 to 30 percent of AI workloads in 2023).

For telcos, this opportunity provides favorable conditions for turning the tide on a decade of diminishing B2B wireline revenue and establishing healthier, more sustainable B2B income streams as they shift from charging monthly usage and capacity to a more lucrative, value-based pricing model. Lumen, for instance, launched ExaSwitch technology to provide customers with a self-service portal where they can configure and connect their edge sites, data centers, and central offices and route traffic without third-party intervention—enabling users to scale capacity up or down based on their needs.[11] In the United Kingdom, British Telecom launched an all-in-one networking and security service for UK corporate and public-sector customers, managed by British Telecom experts and utilizing secure firewall and networking equipment from Fortinet to help manage multisite connections. The company's intelligent network service is designed to help its customers minimize cyber risk, more easily migrate to the cloud, and gain more visibility over their network to manage workloads.[12]

This next generation of SDNs is still evolving. As a result, there isn't yet a single market definition of what features and capabilities are needed or of what software product strategy and business model will work best for different customers. In other words, leaders may need to make significant investments in market research and software development to understand customer needs and expectations and adjust their product offerings and strategy as the market evolves. There are several key questions to consider:

— *Product vision and road map.* What is the vision for an intelligent network? Which capabilities should be prioritized for development in the short, medium, and long term?

— *Build or partner.* Should we develop new capabilities internally or partner with third parties, such as over-the-top players (OTTs), to enhance our offerings?

— *Revenue model and monetization.* Do companies expect these features as natural product improvements, or will they pay for them—and if so, at what price? What are the optimal model and strategy for these capabilities? Will customers buy incremental capabilities in a bundled solution, or should we allow them to select the options they need?

[8] "AI demand drives $5 billion in new business and massive expansion of the internet," Lumen press release, August 5, 2024.
[9] Leasing of dedicated wavelengths on existing fiber optic lines.
[10] "Tame data egress charges in the public cloud," Gartner, September 27, 2023; "Gartner forecasts worldwide public cloud end-user spending to surpass $675 billion in 2024," Gartner press release, May 20, 2024.
[11] "Lumen, Google and Microsoft create new on-demand, optical interconnection ecosystem," Lumen press release, June 12, 2023.
[12] "BT launches a new managed SASE service powered by Fortinet," BT press release, November 25, 2024.

In assessing the opportunity, leaders will need to balance current market uncertainty against the cost of inaction. Network service providers are aggressively developing these advanced services for enterprises, threatening the incremental revenue that telcos currently capture on SDNs.

Path 3: Turning unused space and power into revenue

Often, telcos look to sell unused data center or central office space to investors seeking to convert the space to residential, commercial (including retail and office), or mixed-use developments. A growing opportunity for these underutilized assets is for telco operators to offer the space—either directly or through sale-leaseback agreements—to hyperscalers, colocation providers, GPU-as-a-service (GPUaaS) firms, and large enterprises that need immediate access to data center space and power for their operations.

Such space and power can be extremely valuable in some markets. New data center builds can take upwards of five years, delaying growth and expansion. Even if a company can fast-track construction, power grids are running at capacity in many markets, and energy companies cannot provide power to new builds. That gives telcos an advantage—at least for a few years—in filling the void and capturing revenue from existing assets with minimal investment. For telcos aiming to join the AI compute value chain without substantial risk, transforming and offering unused space and power may provide a practical path, especially for US and European operators that have not yet exited from the data center business as they assess whether separation makes sense.[13] In Asia–Pacific, operators will likely consider this path low-hanging fruit, given their strong presence in the data center business. For instance, many operators in Asia remain national or regional data center operators, with some, like NTT and KDDI, among the top 30 global colocation players.[14]

In some cases, telcos might adopt a colocation business model, serving as landlords and renting the space to tenants with minimal retrofitting, such as installing cages and security cameras for tenants. Operators may also enter into a revenue-sharing agreement, where they might commit to some capital expenditure up front and the "tenant," likely a hyperscaler or GPUaaS provider, brings the business, AI cloud platform, and potentially GPUs.

Verizon's MEC (mobile-edge computing) partnership with AWS uses a hybrid approach to enable enterprises to obtain compute power close to the end user for real-time inferencing. In this model, Verizon provides space and power to AWS, which brings the compute, storage, and customers. But Verizon isn't just a neutral landlord. The company plays an active role in running the AI workloads on its telecommunications network.[15]

Opportunities to turn unused space and power into revenue vary significantly by market based on numerous factors, including the density of data centers in the area and their capacity to support today's intensive AI workloads. In some places with high concentrations[16] of data centers, such as Northern Virginia in the United States, leaders will find less advantageous revenue opportunities than in nontraditional markets where hyperscalers and other market players are looking to break ground.

Where there is demand, telcos should consider what type of data center is needed and how quickly and cost-effectively they can transform an existing space to meet that demand. Is there enough power on-site? Typically, data centers that support AI inferencing at the edge require at least 500 kilowatts of power capacity to be economically attractive.

Are there space constraints that limit potential upgrades? Server racks designed to support

[13] "Can telcos create more value by breaking up?," McKinsey, January 22, 2020.
[14] 451 Research, part of S&P Global Market Intelligence.
[15] "Verizon and AWS team up to deliver 5G edge cloud computing," Verizon press release, December 3, 2019.
[16] "AI power: Expanding data center capacity to meet growing demand," McKinsey, October 29, 2024.

gen AI workloads typically pack more computing power than in traditional data centers, requiring the installation of liquid cooling equipment. Spaces designed to support conventional air-cooling systems often need retrofitting to support these more efficient cooling systems. As telco leaders seek to turn assets into revenue-generating income, prioritizing their efforts for the first few deals around facilities that need little retrofitting or adapting a facility only after they obtain a firm commitment from future tenants may help mitigate their risk.

Path 4: Building a new GPUaaS business

GPUaaS offerings allow organizations to gain remote access to high-performance GPUs hosted in AI-ready data centers without costly up-front investment. GPUaaS providers rent GPU compute using flexible pricing models, such as by the hour (called spot contracts) or for a longer term (reserved contracts). Such offerings can provide access to large-scale GPU clusters for public-sector use cases, including sovereign AI;[17] enterprise training of AI models; or delivery of small GPU clusters close to users for AI inferencing (often sold as inference as a service, or IaaS). This can potentially be an attractive opportunity: Accelerated compute workloads[18] (including for AI and gen AI) are expected to grow at more than 30 percent CAGR and could make up more than two-thirds of data center demand within the next five years (Exhibit 5).

Our research suggests that the addressable GPUaaS market addressed by telcos[19] could range from $35 billion to $70 billion by 2030 globally (Exhibit 6). A majority of the demand is expected to come from North America and Asia. Some leading telcos, such as Verizon in North America, are partnering with GPUaaS players.[20] Others, such as Indosat Ooredoo Hutchinson (IOH), Singtel, and Softbank in Asia, have launched their own GPUaaS offering to market directly, often in collaboration with partners.[21]

This growth opportunity may be especially attractive in countries with regulatory frameworks that support national AI strategies and sovereign AI, given that governments increasingly scrutinize data security risks from non-national providers. One Norwegian telecommunications company, Telenor, for instance, is launching a sovereign AI for the Nordics through its collaboration with NVIDIA.[22] Similarly, Swisscom's new Swiss AI platform enables companies to store and process data in Switzerland.[23]

Operators pursuing this path often use one of two models: either leveraging their data center or renting space from a colocation provider. For telcos that have delayered[24] or sold their data centers, leasing data center space for GPUaaS can be an effective approach to capitalizing on the opportunity without investing in new facilities. However, these telcos will still incur capital costs for GPUs and operating expenses like utilities and maintenance. Our research finds that both models could deliver a return on invested capital ranging from 6 to 14 percent, depending on an operator's current asset base, such as existing data centers in operation and the supply and cost of electricity to its data center location.

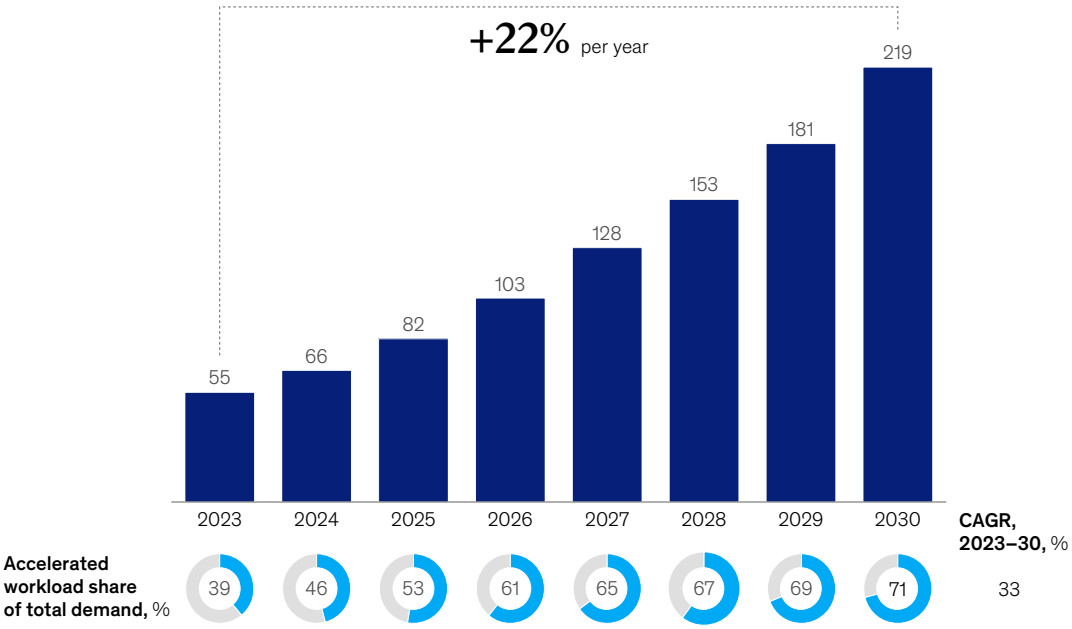
Given the significant capital expenditures to launch a GPUaaS offering (with or without AI-ready data centers), leaders will need to navigate numerous

[17] Sovereign AI refers to the development and deployment of AI systems within a nation's borders to ensure technology independence and data security.
[18] Accelerated workloads are tasks performed on servers equipped with semiconductor accelerators, such as GPUs or ASICs.
[19] Excludes the market share expected to be addressed by hyperscalers.
[20] "Verizon unveils AI strategy to power next-gen AI demands," Verizon press release, January 24, 2025.
[21] "NVIDIA, Indosat plan $200 mln AI centre investment in Indonesia, government says," Reuters, April 4, 2024; "Singtel to introduce GPU-as-a-service, powered by NVIDIA accelerated computing," Singtel press release, March 19, 2024; "NVIDIA and Softbank Corp. accelerate Japan's journey to global AI powerhouse," NVIDIA press release, November 12, 2024.
[22] "Telenor Group announces collaboration with NVIDIA to support its AI-first ambition," Telenor press release, February 26, 2024; "Telenor launches AI factory with HIVE Autonomy as first customer," Telenor press release, November 14, 2024.
[23] Armin Schädeli, "Swisscom launches Swiss AI Platform for companies," Swisscom, November 28, 2024.
[24] Andrea Travasoni, Cristina Dexeus, and Kim Baroudy, "Delayering: An organizational overhaul for growing Europe's telcos," McKinsey, February 23, 2023.

Exhibit 5

Accelerated workloads are expected to be a key driver for 22 percent annual growth in data center demand through 2030.

Estimated global data center demand (midrange scenario), gigawatts



Note: Accelerated workloads are tasks performed on servers equipped with semiconductor accelerators, such as GPUs or ASICs.
Source: McKinsey proprietary datacenter demand model

McKinsey & Company

risks, including uncertain demand, competition from hyperscalers, rapid technology shifts, and potential price drops from increased supply. A lower-risk modular approach could start with a modest investment in AI-ready data centers and GPUs that focuses on securing anchor tenants and building a partner ecosystem. Offering complementary technology services and AI capabilities can increase demand for AI use. Through its services arm, Telefonica Tech, Telefonica launched ten global AI specialist centers with over 400 AI professionals dedicated to researching and developing customer AI use cases. Recently, it announced a gen AI platform with capabilities designed to streamline the development of customizable virtual assistants to improve customer service.[25]

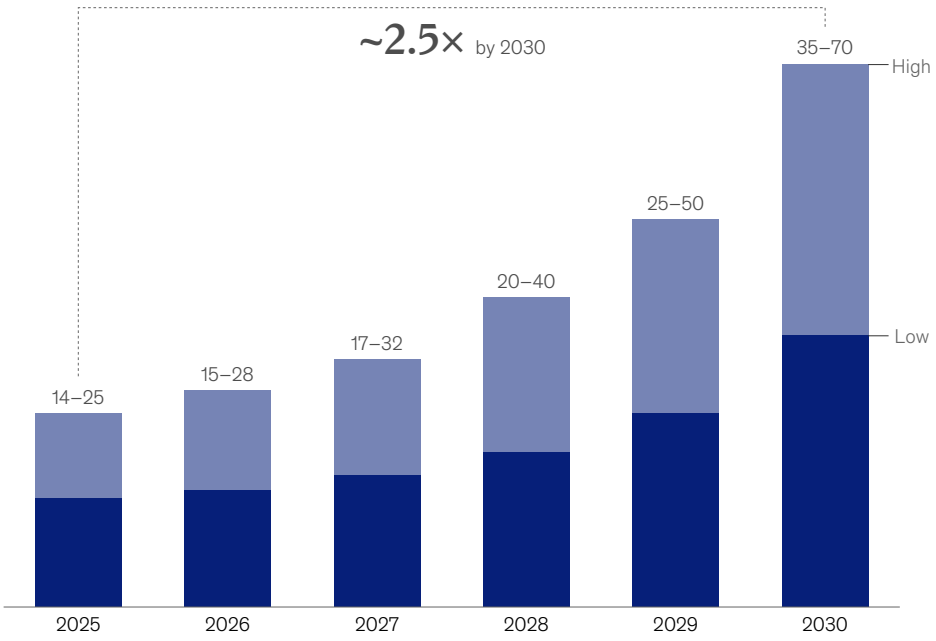
Demand for distributed GPUaaS is also emerging because the development of new data centers is constrained by limited grid power, growth in AI has resulted in high volumes of data throughput, and there is an increasing need for low-latency compute for real-time inferencing. As a result, an emerging opportunity could be to use GPU-based hardware for both RAN and AI workloads (called AI-RAN), which could help optimize network efficiency while simultaneously enabling

[25] "Telefónica Tech accelerates its AI business with ten specialised centres," Telefónica press release, May 30, 2024, "Telefónica Tech launches generative AI platform to create customisable virtual assistants," Telefónica press release, January 28, 2025.

Exhibit 6

Global demand for GPU as a service addressable by telcos is projected to range from $35 billion to $70 billion by 2030.

GPU as a service (GPUaaS) global demand estimate,[1] $ billion



Source: McKinsey proprietary AI demand model

McKinsey & Company

monetization of distributed RAN compute through GPUaaS and IaaS (see sidebar, "The viability of AI-RAN").

Transitioning to a new operating model for success

As leaders identify viable paths, they will likely need to adjust their operating models to fully capitalize on these opportunities. Shifts in four key areas—sales strategies, partnerships, financial evaluations, and communications—are typically needed, with different paths sometimes requiring different operating models.

— *Sales strategy.* In some cases, telcos will need a dedicated sales organization (with exclusive profit and loss) beyond their existing sales team. For example, those offering space and power and data center connectivity will find that hyperscalers, a key customer segment, typically rely on such specialized sales teams to provide fast design and regular cost analysis. To support this, operators will often have to digitize routing data and develop rate cards for dark-fiber pricing to enhance sales responsiveness. In other instances, telcos need to upskill their existing B2B sales teams. For instance, when selling GPUaaS to large enterprises and model developers, B2B sales organizations will need to shift from a traditional solution-focused approach to an advisory-led approach with customer

The viability of AI-RAN

Sharing GPU-based hardware for both RAN and AI workloads (AI-RAN) can potentially transform RAN deployments—and enable new revenues simultaneously. Recognizing this opportunity, in early 2024, key players from the telco and AI value chains formed the AI-RAN Alliance to assess how AI can be embedded in RAN. A central concept of AI-RAN is replacing the baseband units (BBUs) in the mobile network with GPU-based hardware that can be used to run both RAN and AI workloads. In doing so, RAN workloads could be optimized through AI, with preliminary vendor testing indicating improvements in spectral efficiency, network performance, and resource management.[1] With this, telcos could also offer distributed GPUaaS—providing cloud-based GPU access for AI workflows across the mobile network. Here, telcos use existing site locations to address issues of latency, throughput congestion, and energy concentration on the grid for AI computations, as Softbank and T-Mobile are doing to complement centralized compute resources in data centers.[2]

While the merits of AI-RAN are still being tested, three dimensions remain uncertain and need to be resolved as telcos consider its viability:

— *Customer demand.* Who will require distributed AI compute, and for what use cases? Telcos should evaluate potential demand from several types of organizations and usage needs, including hyperscalers distributing workloads beyond centralized locations, enterprises requiring secure local compute resources or redundancy, or use cases currently constrained by throughput or energy availability.

— *Network requirement.* Is the current architecture compatible with the AI-RAN concept? Efficiency of AI-RAN for RAN operations relies heavily on centralization,[3] which is not possible for all telcos due to limitations in expanding fiber networks. Practically, implementation requires synchronization with deployment cycles (for example, a new frequency layer rollout or equipment refresh). Telcos will have to understand whether such an architecture is feasible within their environment and in what time frame.

— *Total cost of ownership.* How will the total cost of ownership for the RAN itself compare with traditional alternatives? The potential additional cost of deploying GPU-based hardware and adapting the network architecture for AI-RAN will have to be weighed against additional revenues from the GPUaaS offering.

Our research suggests the cost-efficiency of AI-RAN (which differs by operator) will determine future evolution. Operators will likely consider three scenarios. First, if AI-RAN proves more cost-effective, telcos will face little downside in adopting it. The financial incentive is clear: Lower costs and the potential to generate new revenue from AI workloads make early adoption a compelling opportunity.

Second, if AI-RAN's total cost of ownership carries only a slight penalty, telcos may consider the additional cost as the price to access future revenue potential, even if they judge the takeoff of distributed GPUaaS as uncertain, provided deployment is practical in their network.

Finally, if AI-RAN comes with a higher price tag, the only players to pursue rollouts will be those that have a competitive edge plus a strong conviction that a new market for GPUaaS is about to open up for telcos. Lacking that, a wait-and-see approach is more natural, but it comes with accepting the risk that someone else may gain an irreversible first-mover advantage in this potentially new AI-compute provisioning landscape.

[1] NVIDIA blog, "Bringing AI-RAN to a telco near you," blog post by Soma Velayutham, October 8, 2024.
[2] "T-Mobile announces technology partnership with NVIDIA, Ericsson and Nokia to advance the future of mobile networking with AI at the center," T-Mobile press release, September 18, 2024; *AI-RAN:Telecom infrastructure for the age of AI,* Softbank, December 2024.
[3] Gerardo de Geest, Gustav Grundin, Ole Jørgen Vetvik, and Nemanja Vucevic, "Telecom networks: Tracking the coming xRAN revolution," McKinsey, February 2023.

success teams providing education, feasibility checks, and pre- and post-sales services with proofs of concept and offerings at scale in key industries. Leveraging existing relationships with anchor tenants or organizations that may commit to a specific capacity (known as capacity uptakers) can also help reduce risk.

— *Partnerships.* New partnerships will often be needed to complement an operator's expertise and accelerate the time to market before a competitor can gain a strong foothold. Teaming with GPU chipmakers, data center providers, data and AI platform vendors, IT service providers, power suppliers, and investors can enable operators to build a full-stack GPUaaS offering to attract customers. Software providers can help operators deliver new intelligent network services for AI more quickly. Systems integrators can provide the managed services necessary to ensure seamless, secure, and efficient connectivity between data centers, allowing operators to focus on other priorities. General contractors specializing in data center retrofitting can enable operators to prepare unused data center space and upgrade power more quickly and cost-effectively than if they tried to do the work themselves.

— *Financial evaluation.* Evaluating these paths can be complex and frequently requires nontraditional approaches. This can include predicting future demand for connecting multiple data centers with fiber infrastructure along the same route instead of focusing on the potential returns from the initial customer. Another way to improve the financial viability and value of fiber deals is negotiating with colocation providers to limit the number of connectivity providers they use—for example, signing deals with only two instead of ten or more. By doing so, operators can achieve higher volumes and greater cost-efficiency, enabling them to lower prices for the enterprises that use these data centers. Those who need to ensure their data centers are "GPU-ready"—including those pursuing space and power and GPUaaS using their own data centers—will want to consider capital and operational costs across five key dimensions: compute, power, cooling, networking, and memory.

— *Communication.* In this highly competitive arena, telco leaders will do well to proactively shape the public narrative about their role within the AI value chain, rather than waiting for investor or customer questions. This includes educating enterprises, investors, and industry analysts about their plans and value-add. AT&T's CEO, for instance, spoke about the implications of AI innovations in a recent interview.[26] Lumen has begun highlighting network KPIs critical for AI, such as optical loss, latency, and capacity, on its website as part of its work to build trust and leadership in the marketplace.[27] Verizon featured its AI strategy prominently during its last earnings call.[28]

With the growing demand for AI infrastructure, telcos have an opportunity to serve as the backbone of the AI era. Success will depend on acting decisively, balancing ambition with pragmatism, and embracing new operating models. Not every path will suit every telco; some may be too risky for certain operators right now. However, the most significant risk may come from inaction, as telcos face the possibility of missing out on their fair share of growth from this latest technological disruption.

[26] "A DeepSeek breakthrough and TikTok's future," *CNBC Squawk Box,* January 27, 2025.
[27] For more information, see Lumen.com.
[28] *4Q 2024 Earnings & Business Update,* Verizon, January 2025.

Abhyuadaya Shrivastava is a partner in McKinsey's Singapore office, where **Lorraine Salazar** is a senior knowledge expert; **Brendan Gaffey** is a senior partner in the Dallas office; **Gustav Grundin** is a partner in the Tokyo office; **Sebastian Cubela** is a partner in the Miami office; **Tomás Lajous** is a senior partner in the New York office, where **Miguel Frade** is an associate partner; **Akshat Agarwal** is a consultant in the Dublin office; **Dapo Orimoloye** is a consultant in the Houston office, and **Nicholas Shaw** is a senior knowledge analyst in the Boston office.

The authors wish to thank Adarsh Chawla, Andrew Copuyoc, Antonius Santoso, Arjita Bhan, Daniel Eisenberg, Ferran Muntaner, Ferry Grijpink, Gerardo de Geest, Hiroyuki Katayama, Ilana Reynolds, Ivfen Foong, Marina Benedet, Pankaj Sachdeva, Rishi Gupta, Satyam Taneja, Shannen Poon, Tatsuya Kabuta, and Wilbur Wang for their contributions to this article.

Copyright © 2025 McKinsey & Company. All rights reserved.

Find more content like this on the
McKinsey Insights App

Scan • Download • Personalize